Pattern Recognition Letters 33 (2012) 934–942

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Lexicon-free handwritten word spotting using character HMMs

Andreas Fischer *, Andreas Keller, Volkmar Frinken, Horst Bunke

University of Bern, Institute of Computer Science and Applied Mathematics, Neubrückstrasse 10, 3012 Bern, Switzerland

ARTICLE INFO

Article history:
Available online 10 September 2011

Keywords:
Handwriting recognition
Keyword spotting
Hidden Markov Models

ABSTRACT

For retrieving keywords from scanned handwritten documents, we present a word spotting system that is based on character Hidden Markov Models. In an efficient lexicon-free approach, arbitrary keywords can be spotted without pre-segmenting text lines into words. For a multi-writer scenario on the IAM off-line database as well as for two single writer scenarios on historical data sets, it is shown that the proposed learning-based system outperforms a standard template matching method.

1. Introduction

Handwriting recognition of scanned or photographed text images is still a widely unsolved problem in pattern recognition, although it has been an active area of research for several decades (Vinciarelli, 2002). The automatic recognition of handwritten text images is an *offline* task that is considered to be harder than *online* recognition, where temporal information can be exploited (Plamondon and Srihari, 2000). For large vocabularies and different writing styles in general (Bunke and Varga, 2007), and for degraded historical manuscripts in particular (Antonacopoulos and Downton, 2007), the accuracy of an automatic transcription is far from being perfect. Under these conditions, word spotting has been proposed instead of a complete transcription for the restricted task of retrieving keywords from document images (Manmatha and Croft, 1997).

Handwritten word spotting is of great interest in different application areas. For modern handwriting, an important application is given by automatic mail sorting. Nowadays, large companies still receive a high volume of handwritten correspondence. One might be interested, for example, to give more priority to mails containing the word "urgent" (Rodriguez and Perronnin, 2009). For historical documents, a key application is given by integrating handwritten documents in digital libraries (Nagy and Lopresti, 2006). In the context of cultural heritage preservation, many libraries all around the world have digitized their most valuable old handwritings, ranging from religious Old Greek manuscripts to handwritings from Modern Ages, e.g., George Washington's papers at the Library of Congress. Word spotting can be used for indexing the vast amount of available document images, in order to make them amenable to searching and browsing (Rath et al., 2004).

Two different approaches to handwritten word spotting can be distinguished in the literature. On the one hand, *template-based* methods (Manmatha et al., 1996; Zhang et al., 2004; Bhardwaj et al., 2008; Rothfeder et al., 2003; Rath and Manmatha, 2003, 2007; Leydier et al., 2009; Adamek et al., 2007; Rodriguez and Perronnin, 2008; Terasawa and Tanaka, 2009) match a query word image with labeled keyword template images. An advantage of this approach is that template images are rather easy to obtain, even if the underlying language and its alphabet are unknown. However, such systems are limited by the fact that for each possible keyword that is to be spotted, at least one template image is needed and unknown out-of-vocabulary words cannot be spotted at all. Also, they typically have a low generalization capability to unknown writing styles.

On the other hand, *learning-based* methods (Rodriguez and Perronnin, 2009; Choisy, 2007; Perronnin and Rodriguez-Serrano, 2009; Edwards et al., 2004; Chan et al., 2006; El Yacoubi et al., 2002; Thomas et al., 2010; Frinken et al., 2010) employ statistical learning methods to train a keyword model that is then used to score query images. A very general approach was recently introduced in (Rodriguez and Perronnin, 2009), where the learning-based approach is applied at word level based on Hidden Markov Models (HMMs). The trained word models are expected to show a better generalization capability than template images. However, the word models still need a considerable amount of templates for training and the system is not able to spot out-of-vocabulary keywords.

When the learning-based approach is applied at character level, a word spotting system obtains, in principle, the capability to spot arbitrary keywords by concatenating the character models appropriately. Although this approach is well-known for the recognition of speech (Rose and Paul, 1990) and poorly printed documents (Chen et al., 1993), only few reports can be found in the handwriting recognition literature. While an earlier approach presented in (Edwards et al., 2004) was based on a small set of manually extracted letter-templates, there is a recent tendency towards fully-fledged learning-based systems at character level (Thomas et al., 2010; Frinken et al., 2010).

* Corresponding author. Fax: +41 31 631 32 62.
E-mail address: afischer@iam.unibe.ch (A. Fischer).

0167-8655/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2011.09.009

In the present paper, we propose a learning-based word spotting system for unconstrained handwritten text that is based on character HMMs. When compared with previous word spotting systems, the proposed approach has several advantages. First, the system is segmentation-free at the training stage as well as at the recognition stage, i.e., it is not dependent on a segmentation of text lines into words that can be prone to errors. An advantage over learning-based systems at word level is given by the fact that only a small number of character classes needs to be trained for which a rather large number of training samples is available in a given handwritten text. Furthermore, the proposed system is able to spot arbitrary keywords that are not required to be known at the training stage. When compared with transcription-based systems, the proposed word spotting approach has the advantage that no lexicon is used. For a large vocabulary task, such a lexicon imposes a high computational complexity.

A known disadvantage of the proposed system is, however, that it requires a set of transcribed text line images for training that may be costly to obtain, in particular for historical documents. If neither the language nor the alphabet of a historical document are known, template-based image matching might be the only option available.

In an experimental evaluation, we have tested the proposed system on three data sets that represent different word spotting conditions. First, the system was tested on the IAM off-line database that contains modern English texts written by several hundred writers. The main difficulty with this database is to spot keywords in unknown handwriting styles, taking into account a large amount of training data from other writers. Secondly, tests were conducted on the George Washington database that includes a small collection of handwritten historical letters sharing a common writing style. Here, the main difficulty is to cope with the small amount of training data available. Finally, the system was tested on the Parzival database that contains a larger collection of medieval manuscripts with a common writing style. Taking into account a rather large amount of training data, the best word spotting conditions are met for this database.

The proposed system is compared with a well-established template matching method based on Dynamic Time Warping (DTW) (Rath and Manmatha, 2007). It is demonstrated that the proposed learning-based approach outperforms the reference system not only for the multi-writer case, as expected, but also for both single-writer scenarios.

The remainder of the paper is organized as follows. In Section 2, the proposed word spotting system is presented in detail. Section 3 introduces the reference system that is used for experimental evaluation. Next, Section 4 describes the experimental setup. In particular, the data sets are discussed in Sections 4.1–4.3. Results and discussion are given in Section 5 and finally, conclusions are drawn in Section 6.

1.1. Related work

Keyword spotting has been applied to speech (Cory et al., 1980; Rose and Paul, 1990) and poorly printed documents (Chen et al., 1993; Kuo and Agazzi, 1994) before. For handwritten text, it was proposed a few years later in (Manmatha et al., 1996). In a template-based approach, the Scott and Longuet-Higgins distance (SLH) (Scott and Longuet-Higgins, 1991) was used in (Manmatha et al., 1996) to compare keyword template images and unknown word images. In the following, other features based on global image characteristics have been proposed, e.g., gradient, structural and convexity features (GSC) (Zhang et al., 2004), and features based on moments of binary images (Bhardwaj et al., 2008).

A different template-based approach is given by using local features in conjunction with elastic matching. Notable work in this domain includes the corner features proposed in (Rothfeder et al., 2003), sliding window features in combination with Dynamic Time Warping (DTW) (Rath and Manmatha, 2003), and gradient angle features in combination with a cohesive elastic distance (Leydier et al., 2009). DTW, in particular, is well-established in the field and has been used in combination with different features in recent work, e.g., based on word profiles (Rath and Manmatha, 2007), closed contours (Adamek et al., 2007), and local gradients (Rodriguez and Perronnin, 2008; Terasawa and Tanaka, 2009). In (Terasawa and Tanaka, 2009), an extension of DTW is presented that allows to match keyword templates with complete text lines rather than segmented word images.

A very general learning-based approach at word level was presented in (Rodriguez and Perronnin, 2009). Based on local gradient features, posterior probabilities of keyword HMMs are used for keyword spotting in conjunction with universal vocabularies for score normalization. A similar approach was presented in (Choisy, 2007) for non-symmetric half plane HMMs (NSHP-HMMs). Instead of using posterior probabilities, a scoring method based on Fisher kernels has been proposed in (Perronnin and Rodriguez-Serrano, 2009).

For spotting arbitrary keywords, the learning-based approach has been investigated at character level as well. Character template images were used in (Edwards et al., 2004; Chan et al., 2006) to train generalized Hidden Markov Models (gHMMs) for keyword spotting. While promising results were reported for historical Latin manuscripts (Edwards et al., 2004) and Arabic scripts (Chan et al., 2006), an automatic acquisition of such character templates from handwritten text images is difficult in general. The same limitation is faced in (Cao and Govindaraju, 2007), where segmented character images and Gabor features were used in a template-based method for keyword spotting.

In an earlier work (El Yacoubi et al., 2002), character HMMs were used to spot street names for the constrained task of address reading. In a segmentation-free approach, character models are trained on complete text line images and are then connected to model keywords as well as general non-keyword text. Recently, this line of research was followed in (Thomas et al., 2010) for unconstrained handwritten word spotting. Another system that is based on learning at character level was presented in (Frinken et al., 2010) based on bidirectional long short-term memory neural networks (BLSTM-NNs).

In contrast to El Yacoubi et al. (2002) and Thomas et al. (2010), the method proposed in this paper is not dependent on a lexicon that imposes a high computational complexity for large vocabulary tasks. Instead of a comparison with other lexicon words, a log -odds scoring with respect to a general filler model is employed as a confidence measure for keyword spotting. This technique is well-known from other application domains of HMMs, e.g., speech recognition (Rose and Paul, 1990) and bioinformatics (Barrett et al., 1997), and has been used similarly in (Rodriguez and Perronnin, 2009).

1.2. Contribution

In this work, a novel lexicon-free keyword spotting system using trained character HMMs is presented for handwritten word spotting. On several data sets, its superior performance over a standard template matching approach is demonstrated.

This article is an extended version of a conference paper (Fischer et al., 2010a) that has introduced a preliminary version of the system. In general, a more thorough description and evaluation of the system is provided in this work. In particular, in Section 2.4, a theoretical justification for the proposed text line scoring for keyword spotting is given. The computational complexity of this scoring mechanism is analyzed in Section 2.5.